A NEW ALGORITHM FOR ITERATIVE DECONVOLUTION OF SPARSE SPIKE TRAINS

Y. Goussard, G. Demoment and J. Idier

Laboratoire des Signaux et Systèmes (CNRS / ESE / UPS)
École Supérieure d'Électricité, Plateau de Moulon
91192 Gif-sur-Yvette Cédex, France.

ABSTRACT

This article presents an new iterative algorithm for deconvolution of Bernoulli-Gaussian processes. This detection-estimation problem is formulated as that of a change of initial conditions in linear least-squares estimation. An algorithm with a very simple structure is obtained. It allows the evaluation of either marginal or joint likelihood criteria without any approximation; the resulting method is easy to implement, computationally inexpensive and remains nearly optimal.

I. INTRODUCTION

The subject of this communication is the restoration of sparse spike trains distorted by a linear system and corrupted by an additive observation noise. This problems arises in many areas of applied physics, e.g., seismic exploration, non-destructive evaluation, medical ultrasonic imaging, or more generally when one wishes to characterize a non-homogeneous propagation medium from measurements performed at its boundaries. Under reasonable assumptions, the observations can be considered as the convolution product of two signals: the wavelet, i.e., the incident waveshape, and the reflectivity or first derivative of the acoustic impedance which characterizes the unknown medium. The problem consists of restoring the reflectivity from the observed data and from any available information on the wavelet and on the observation noise. In general, this is a *myopic deconvolution* problem since neither the wavelet nor the statistical properties of the noise are known precisely. Here, the scope of the study is limited to *simple deconvolution* in which the only unknown quantity is the reflectivity.

Here, we are interested in the characterization of stratified media with homogeneous layers. In this case, the reflectivity vanishes everywhere except at the boundaries between layers, and thus appears as a pulse process. Simple deconvolution of such signals presents the usual difficulties of the resolution of ill-posed problems and some additional ones caused by the particular shape of the reflectivity and by specific implementation constraints which depend on the application domain. Methods based upon L_p norms [1], minimum entropy criteria [2] or multi-pulse coding techniques [3] have been proposed. In these methods, neither the specific characteristics of the reflectivity nor the ill-posed nature of the deconvolution problem are explicitly accounted for. A Bayesian approach was introduced by Mendel *et al.* [4]-[6] and followed by others [7]-[9]. It is based upon a Bernoulli-Gaussian (B-G) description of the reflectivity and maximization of likelihood-type criteria. Deconvolution of B-G signals is a detection-estimation problem which cannot be solved exactly due to the usual size of the signals to be processed. Therefore all methods of this type are sub-optimal to some extent. Some of them exhibit a globally recursive structure suited to on-line data processing. Their drawback is a lack of robustness, notably when the frequency content of the wavelet is poor. Iterative methods do not present the same disadvantage. On the other hand, they are generally rather computationally demanding, which makes them complicated to implement and to use.

The method presented here is aimed at partly correcting this drawback. It is an iterative procedure based upon a finite impulse response (FIR) representation of the wavelet. The derivations presented below are closely related to those of Mendel *et al.* This is not surprising since the models used for the reflectivity and for the noise, as well as the criteria which are maximized, are almost identical. However, a very simple algorithmic structure, which to our knowledge has not been pointed out and exploited before, is brought to evidence here. It yields considerable simplification of the implementation of the method and greatly decreases its computational requirements.

II. PROBLEM FORMULATION

Under the assumptions that all phenomena are linear and that the observation noise is additive, the input-output equation of the system can be written as

$$z(k) = \sum_{i=0}^{n} h(i)\, x(k-i) + n(k) \qquad 1 \le k \le P \tag{1}$$

where z, x and n denote the observations, the unknown reflectivity and the observation noise, respectively. For the

CH2847-2/90/0000-1547 $1.00 © 1990 IEEE

sake of simplicity, wavelet h is assumed to be time-invariant with a support region of $n+1$ samples. Concatenation of the samples of z, n and x into vectors $\boldsymbol{z}$, $\boldsymbol{n}$ and $\boldsymbol{x}$ of respective dimensions P, P and N allows (1) to be rewritten in the matrix form

$$\boldsymbol{z} = \boldsymbol{H}\,\boldsymbol{x} + \boldsymbol{n} \tag{2}$$

where $\boldsymbol{H}$ contains the shifted samples of wavelet h. Observation noise n is assumed to be Gaussian white centered with variance r^n and independent from signal x. The Bernoulli-Gaussian model of the input can be expressed as

$$\begin{aligned} &t(k)\text{: Bernoulli random variable (RV)} \begin{cases} P\{t(k)=1\} = \lambda \\ P\{t(k)=0\} = 1-\lambda \end{cases} \\ &x(k)\text{: Centered Gaussian RV with variance } r^x\,t(k)) \\ &x,\ t\text{: White processes} \end{aligned} \tag{3}$$

When λ is small with respect to 1, (3) does define a sparse spike train. It should be noticed that conditionally to the knowledge of Bernoulli sequence t, x is a non-stationary Gaussian process. Unlike Mendel *et al.* [4], [5], no product model is used in the definition of the B-G process (3). This simplifies the estimation problem, with the beneficial effect of an improved robustness of the maximum-likelihood estimators [10].

As mentioned earlier, restoration of a B-G process requires two operations: Detection of the Bernoulli sequence t and estimation of the amplitudes of x. In a maximum-likelihood approach, two strategies can be used: In the first one, detection and estimation are carried out *simultaneously* with the maximization of the joint likelihood function

$$J_J(\boldsymbol{t}, \boldsymbol{x}) \triangleq p\{\boldsymbol{t}, \boldsymbol{x}|\, \boldsymbol{z}\} \propto p\{\boldsymbol{z}|\, \boldsymbol{t}, \boldsymbol{x}\}\; p\{\boldsymbol{x}|\, \boldsymbol{t}\}\; P\{\boldsymbol{t}\} \tag{4}$$

with respect to $\boldsymbol{t}$ and $\boldsymbol{x}$. However, it has been shown [5], [6] that such a *simultaneous* detection-estimation operation can be performed sequentially in a two step procedure: First, detection of the Bernoulli sequence through maximization of an appropriate criterion J_D, which yields an estimate $\hat{\boldsymbol{t}}$ of $\boldsymbol{t}$, and second, estimation of the amplitudes through maximization of $J_J(\hat{\boldsymbol{t}}, \boldsymbol{x})$ with respect to $\boldsymbol{x}$.

The second strategy has a built-in sequential structure: The detection criterion used to obtain the estimate $\hat{\boldsymbol{t}}$ of Bernoulli sequence is the marginal likelihood function

$$J_M(\boldsymbol{t}) \triangleq p\{\boldsymbol{t}|\, \boldsymbol{z}\} \propto p\{\boldsymbol{z}|\, \boldsymbol{t}\}\; P\{\boldsymbol{t}\} \tag{5}$$

Then, the amplitude of the reflectivity is estimated through maximization of J_J exactly in the same way as in the *simultaneous* approach. When $\boldsymbol{t}$ is known, J_J simply represents the *a posteriori* likelihood of $\boldsymbol{x}$. Since all phenomena are linear, the normality of the conditional distribution of $\boldsymbol{x}$ allows us to estimate the amplitudes with the classical maximum *a posteriori* formulas in a linear and Gaussian setting (primed quantities are transposed)

$$\hat{\boldsymbol{x}} = \boldsymbol{\Pi}\boldsymbol{H}'\,\boldsymbol{B}^{-1}\,\boldsymbol{z} \tag{6a}$$

$$\boldsymbol{B} = \boldsymbol{H}\,\boldsymbol{\Pi}\,\boldsymbol{H}' + r^n\,\boldsymbol{I} \tag{6b}$$

In (6), $\boldsymbol{\Pi}$ denotes the *a priori* covariance matrix of $\boldsymbol{x}|\boldsymbol{t}$. It follows directly from the definition of a B-G process that

$$\boldsymbol{\Pi} = r^x\,\boldsymbol{T} \quad \text{with} \quad \boldsymbol{T} \triangleq \text{Diag}\{t(k)\}_{1\le k\le N} \tag{7}$$

Since the estimation step is the same in both strategies, the difference only comes from the detection operation. In other words, the different expressions of the detection criteria reflect the intrinsic differences of the two strategies. When a joint likelihood is used, the detection criterion can be shown to take the following logarithmic form [5], [6], [11]

$$L_J(\boldsymbol{t}) = -\boldsymbol{z}'\,\boldsymbol{B}^{-1}\,\boldsymbol{z} - N_e \ln\left(2\pi r^x\right) - 2N_e \ln\left(\frac{1-\lambda}{\lambda}\right) \tag{8}$$

where N_e denotes the number of non-zero samples of the Bernoulli sequence. Derivation of (8) makes use of the identity

$$\frac{\left(\boldsymbol{z} - \boldsymbol{H}\,\hat{\boldsymbol{x}}\right)'\left(\boldsymbol{z} - \boldsymbol{H}\,\hat{\boldsymbol{x}}\right)}{r^n} + \frac{\hat{\boldsymbol{x}}'\,\boldsymbol{T}\,\hat{\boldsymbol{x}}}{r^x} = \boldsymbol{z}'\,\boldsymbol{B}^{-1}\,\boldsymbol{z} \tag{9}$$

where $\hat{\boldsymbol{x}}$ is given by (6). When a marginal likelihood is used, the expression of the detection criterion follows directly from (5) and takes the following logarithmic form

$$L_M(\boldsymbol{t}) = -\boldsymbol{z}'\,\boldsymbol{B}^{-1}\,\boldsymbol{z} - \ln(|\boldsymbol{B}|) - 2N_e \ln\left(\frac{1-\lambda}{\lambda}\right) \tag{10}$$

Since the detection criteria have very close expression, we define a unique notation

$$L(\boldsymbol{t}) \triangleq -\boldsymbol{z}'\,\boldsymbol{B}^{-1}\,\boldsymbol{z} + \varphi(\boldsymbol{t}) \tag{11}$$

where the respective expressions of $\varphi(\boldsymbol{t})$ in the joint and marginal cases can be easily deduced from (8) and (10). The similarity of the expressions of the detection criteria, the linear structure of the amplitude estimator (6), and the fact that the Bernoulli sequence affects $\hat{\boldsymbol{x}}$ only as an initial condition through the *a priori* covariance matrix $\boldsymbol{\Pi}$, are the key points for the derivation of the restoration algorithm.

III. DERIVATION OF THE ALGORITHM

Exact maximization of $L(\boldsymbol{t})$ would require its evaluation for the 2^N possible realizations of the Bernoulli sequence. As mentioned earlier, it is unreasonable to perform the corresponding computations for realistic values of N. It is more reasonable to explore only a subset of the possible realizations of $\boldsymbol{t}$, and to avoid a complete evaluation of $L(\boldsymbol{t})$ by a direct application of (8) or (10). For this purpose, the notion of *neighboring Bernoulli sequences* is defined as follows: Two sequences are neighbors if all their coordinates are identical, except one at most. Then, the set of possible realizations of $\boldsymbol{t}$ is explored by "jumping" from neighbor to neighbor. Such a procedure is interesting only if simple relationships exist between the criterion values of two neighboring Bernoulli

sequences. This point is critical to the efficiency of the restoration method in practical situations.

Finally, the way in which the possible realizations of $\boldsymbol{t}$ are scanned needs to be determined. One of the simplest ways consists of of selecting the Bernoulli sequence which maximizes L over the whole neighborhood of an initial sequence t_0 and of repeating the procedure until a local maximum is reached. This is exactly the technique proposed by Kormylo & Mendel [4] with the SMLR detector. Its major drawback is that convergence to the global maximum is not guaranteed. In order to avoid this difficulty, other deterministic or stochastic maximization methods could be used, but this point will not be pursued here. We now derive equations which link the criterion values of two neighboring sequences.

III-1. Basic algorithm.

In this section, 0 and k respectively index quantities related to the initial Bernoulli sequence t_0 and to the neighboring sequence t_k which differs from t_0 by its kth coordinate. Let v_k denote the N-dimensional vector whose coordinates are all zero except the kth one which is equal to 1. We also define the auxiliary quantities

$$\boldsymbol{A} \triangleq \boldsymbol{H}' \boldsymbol{B}^{-1} \boldsymbol{H} \tag{12}$$

$$\boldsymbol{w} \triangleq \boldsymbol{H}' \boldsymbol{B}^{-1} z \tag{13}$$

$$\rho_k \triangleq \varepsilon_k (r^x)^{-1} + v_k' \boldsymbol{A}_0 v_k \tag{14}$$

where the ε_k takes the value 1 (resp. –1) when a 1 is added to (resp. removed from) sequence t_0. We now seek a relationship between $L(t_k)$ and $L(t_0)$. It appears clearly in (11) that matrix $\boldsymbol{B}$ plays a central role. Starting from (7), we have

$$\Pi_k = \Pi_0 + \varepsilon_k v_k r^x v_k' \tag{15}$$

Then, using (6b) and applying the matrix inversion lemma to full rank matrix $\boldsymbol{B}_k$ yields

$$\boldsymbol{B}_k^{-1} = \boldsymbol{B}_0^{-1} - \boldsymbol{B}_0^{-1} \boldsymbol{H} v_k \rho_k^{-1} v_k' \boldsymbol{H}' \boldsymbol{B}_0^{-1} \tag{16}$$

From (16), we immediately deduce that

$$\boldsymbol{w}_k = \boldsymbol{w}_0 - \boldsymbol{A}_0 v_k \rho_k^{-1} v_k' \boldsymbol{w}_0 \tag{17}$$

$$\boldsymbol{A}_k = \boldsymbol{A}_0 - \boldsymbol{A}_0 v_k \rho_k^{-1} v_k' \boldsymbol{A}_0 \tag{18}$$

$$z' \boldsymbol{B}_k^{-1} z = z' \boldsymbol{B}_0^{-1} z - \boldsymbol{w}_0' v_k \rho_k^{-1} v_k' \boldsymbol{w}_0 \tag{19}$$

And we obtain the following algorithm

$$\boldsymbol{w}_k = \boldsymbol{w}_0 - \boldsymbol{k}_k \rho_k^{-1} v_k' \boldsymbol{w}_0 \tag{20a}$$

$$\boldsymbol{k}_k = \boldsymbol{A}_0 v_k \tag{20b}$$

$$\rho_k = \varepsilon_k (r^x)^{-1} + v_k' \boldsymbol{k}_k \tag{20c}$$

$$\boldsymbol{A}_k = \boldsymbol{A}_0 - \boldsymbol{A}_0 v_k \rho_k^{-1} v_k' \boldsymbol{A}_0 \tag{20d}$$

$$L(t_k) = L(t_0) + \boldsymbol{w}_0' v_k \rho_k^{-1} v_k' \boldsymbol{w}_0 + \varphi(t_k) - \varphi(t_0) \tag{20e}$$

For the algorithm to be completely defined, it is necessary to derive the exact expression of $\varphi(t_k) - \varphi(t_0)$ for both joint and marginal likelihood functions. Before proceeding to that point, it should be noted that the first four equations of (20) exhibit the familiar structure of a recursive least-squares algorithm. Moreover, most of the computations occur when $\boldsymbol{w}$ and $\boldsymbol{A}$ are updated with (20a) and (20d), and such events happen only once per iteration, when a sequence t_k is selected as a new starting point. Since v_k has a very sparse structure, the other equations are not computationally demanding. Therefore, (20) appears as a very simple and easy to implement basic cell of a SMLR detector, provided that $\varphi(t_k) - \varphi(t_0)$ can be evaluated easily. We now examine this point.

When a joint likelihood function is used, it follows directly from (8) that

$$\varphi_J(t_k) - \varphi_J(t_0) = -\varepsilon_k \left(\ln(2\pi r^x) + 2 \ln\left(\frac{1-\lambda}{\lambda}\right) \right) \tag{21}$$

and since r^x and λ are known quantities, $\varphi_J(t_k) - \varphi_J(t_0)$ can be evaluated to within its sign during the initialization phase.

For the marginal likelihood, we have

$$\varphi_M(t_k) - \varphi_M(t_0) = -\varepsilon_k \left(\ln\left(\frac{|\boldsymbol{B}_k|}{|\boldsymbol{B}_0|}\right) + 2 \ln\left(\frac{1-\lambda}{\lambda}\right) \right) \tag{22}$$

It can be shown from (16) that [11]

$$|\boldsymbol{B}_k| = \varepsilon_k r^x \rho_k |\boldsymbol{B}_0| \tag{23}$$

which yields

$$\varphi_M(t_k) - \varphi_M(t_0) = -\varepsilon_k \left(\ln(\varepsilon_k r^x \rho_k) + 2 \ln\left(\frac{1-\lambda}{\lambda}\right) \right) \tag{24}$$

Therefore, for both criteria, evaluation of $\varphi(t_k) - \varphi(t_0)$ does not present any difficulty. It should finally be checked that the amplitude estimates can be computed easily for a given Bernoulli sequence. From (6a) and the definition of $\boldsymbol{w}$, we immediately obtain

$$\hat{\boldsymbol{x}} = \Pi \boldsymbol{w} \tag{25}$$

Algorithm (20), and formulas (22), (24) and (25) show that simple relationships exist between the criterion values of two neighboring sequences in both joint and marginal cases, and that the estimate of the amplitudes can be recovered easily for any Bernoulli sequence.

III-2. Iterative procedure.

As indicated earlier, the restoration procedure consists of selecting the Bernoulli sequence which maximizes the detection criterion over the whole neighborhood of an initial sequence, and of repeating the procedure until convergence is observed. This guaranties an increase of the criterion value at each iteration. Since the number of possible realizations of the Bernoulli sequence is finite, the procedure necessarily converges in a finite — but perhaps very large — number of iterations. The structure of the restoration procedure, as well as an evaluation of the numerical cost, are indicated below. The figures relative to the initialization phase correspond to a uniformly zero Bernoulli sequence, and the computation of constant quantities used in (21) and (24) is not taken into

account. For the sake of simplicity, the number of significant samples of $\boldsymbol{x}$ and $\boldsymbol{z}$ is assumed to be identical and equal to N.

(1) INITIALIZATION	
Specification of t_0, z, $\boldsymbol{H}$, r^n, r^x, λ	0 mult.
Computation of $\boldsymbol{B}_0^{-1}$	0 mult.
Computation of $\boldsymbol{A}_0$	$n^3 + 1$ mult.
Computation of $\boldsymbol{w}_0$	$N\,(n+1)$ mult.
Computation of $L_J(t_0)$	$N + 1$ mult.
Computation of $L_M(t_0)$	$N + 3$ mult.
(2) ITERATION — For $k \in [1, N]$	
Update of $t(k)$	0 mult.
Computation of $\boldsymbol{k}_k$	0 mult.
Computation of ρ_k	0 mult.
Computation of $\boldsymbol{w}_0\, v_k$	0 mult.
Computation of $L_J(t_k)$	2 mult.
Computation of $L_M(t_k)$	4 mult.
(3) CONVERGENCE TEST	
Selection of $L(t_1) = \mathrm{Max}\{L(t_k)\}$, $1 \le k \le N$	0 mult.
IF $L(t_1) \le L(t_0)$ THEN	
Computation of $\hat{\boldsymbol{x}}$	N_e mult.
CONVERGENCE	
ELSE	
Update of t_0	0 mult.
Update of $L(t_0)$	0 mult.
Update of $\boldsymbol{w}_0$	N+1 mult.
Update of $\boldsymbol{A}_0$	$N(N+3)$ / 2 mult.
RETURN TO (2)	

It appears that the numerical cost of the procedure is rather low, and is almost identical for the joint and marginal likelihood functions. The largest part of the computations is performed during the iterative search of a local maximum of the detection criterion. The complete test of a whole neighborhood requires $O(N^2)$ multiplications, and such a figure allows real-size signals to be processed on small workstation-type computers. The numerical efficiency of the algorithm is mainly the consequence of two points: An appropriate choice of auxiliary variables ($\boldsymbol{w}$ and $\boldsymbol{B}$), and a direct use of matrix $\boldsymbol{H}$ which is made easier by the FIR representation of the wavelet. It should be stressed that the procedure can be easily extended to the case of a time-varying wavelet: $\boldsymbol{H}$ just needs to be initialized according to the known variations of h. Furthermore, since $\boldsymbol{H}$ is used explicitly only in the initialization step of the procedure, the modification leaves the basic algorithm (20) and the iterative optimization procedure unchanged. The major drawback of the method is that matrix $\boldsymbol{A}$ of dimensions (N, N) needs to be stored and manipulated. However, storage of all variables in central memory does not present any difficulty on workstation-type computers.

Our practical experience of the method confirms its ease of implementation and its numerical efficiency. Results obtained on synthetic data are in agreement with what has been previously reported about the behavior of maximum-likelihood deconvolution of B-G processes[4]-[6], [11]: It is preferable to use the marginal likelihood estimator rather than the joint likelihood estimator, and the iterative optimization procedure generally converges to a meaningful local maximum.

IV. CONCLUSION

In this communication, a new algorithm for iterative deconvolution of B-G processes has been presented. It exhibits the structure of a recursive least-squares algorithm, which makes it easy to implement. A low numerical count is achieved, thanks to an appropriate choice of auxiliary variables. Important points for obtaining such characteristics are the FIR representation of the wavelet, and of course the Gaussian distribution of $\boldsymbol{x}$ conditionally to the knowledge of the Bernoulli sequence. Hence, the method appears to be an interesting alternative to existing restoration techniques.

V. REFERENCES

[1] R. Yarlagadda, J.B. Bednar and T.L. Watt, "Fast Algorithms for Lp Deconvolution," *IEEE Trans. Acoustics, Speech & Signal Processing*, **ASSP-33,** 174-182, 1985.

[2] R.A. Wiggins, "Minimum Entropy Deconvolution," *Proc. IEEE Intern. Symp. Computer-Aided Seismic Analysis & Discrimination*, Falmouth, MA, 7-14, 1977.

[3] Y. Goussard, G. Demoment and Y. Grenier, "Déconvolution de processus impulsionnels par algorithmes rapides: Approches AR et MA," *Proc. 11° Colloque GRETSI*, **761-764,** Nice, France, 1987.

[4] J. Kormylo and J.M. Mendel, "Maximum-likelihood detection and estimation of Bernoulli-Gaussian processes," *IEEE Trans. Information Theory*, **IT-28,** 482-488, 1982.

[5] J.M. Mendel, "Optimal seismic deconvolution," Academic Press, New York, 1983.

[6] J. Goutsias and J.M. Mendel, "Maximum-likelihood deconvolution: An optimization theory perspective," *Geophysics*, **51,** 1206-1220, 1986.

[7] A.K. Mahalanabis, S. Prasad and K.P. Mohandas, "Recursive decision-directed estimation of reflection coefficients for seismic data deconvolution," *Automatica*, **18,** 721-726, 1982.

[8] S.D. Kollias and C.C. Halkias, "An instrumental variable approach to minimum-variance seismic deconvolution," *IEEE Trans. Geoscience & Remote Sensing*, **GE-23,** 778-788, 1985.

[9] Y. Goussard and G. Demoment, "Recursive deconvolution of Bernoulli-Gaussian processes using a MA representation," *IEEE Trans. Geoscience & Remote Sensing*, **GE-27,** 384-394, 1989.

[10] F. Monfront, "Déconvolution de signaux Bernoulli-gaussiens par maximum de vraisemblance," Rapport de stage de DEA, Université de Paris-Sud, Centre d'Orsay, 1988.

[11] Y. Goussard, "Deconvolution de processus aléatoires non-gaussiens par maximisation de vraisemblances," Thèse de Doctorat, Université de Paris-Sud, Centre d'Orsay, 1989.